ORIGINAL RESEARCH PAPER

Pharmaceutical Science

BIOSTATISTICAL SOFTWARE: MICROSOFT EXCEL, SPSS AND MINITAB MAKES FINDING CENTRAL TENDENCY AND DISPERSION EASY AND EFFECTIVE

KEY WORDS: Biostatistical software, Dispersion, Central tendency, Microsoft excel, SPSS, Minitab

Tapan Kumar Mahato*

B.Pharmacy College Rampura, At Rampura, PO Kakanpur, Taluka Godhra, District Panchmahal, Gujarat 388713, India*Corresponding Author

ABSTRACT

When we were in school, we were taught how to use a note book and a pen to determine central tendency and dispersion. The entire process took too long and required too much time. These days, we can accomplish this using a programme called biostatistical software. These programmes reduce labour costs as well as time. India consistently ranks as one of the top five pharmaceutical producing nations in the world. The use of software, which allows a work to be completed with less effort and more accuracy, is growing steadily. In research, the use of computers and innovative software is becoming more and more frequent, particularly in the field of biostatistics, which deals with the application of mathematics to biological systems (living things). Due to its usage in both experimental and observational research, biostatistics is advantageous not only to the pharmaceutical business but also to medical and paramedical schools (such as nursing and pharmacy schools). Pen and paper calculations and measurements take an enormous amount of time and are inefficient. Currently, a variety of statistical software is available for the fulfilment of tasks that can quickly produce results. Software programmes used for statistical analysis include Minitab, SPSS, R studio and Microsoft Excel. This article will go over various biostatistical software programmes and how to utilise them to calculate central tendency and dispersion. Because it is free, versatile and simple to use MS Excel is the biostatistical programme that is utilised the most. Cons and benefits can be found in every piece of software. Microsoft Excel is free, whereas other programmes require a monthly or annual fee. There are available demo versions of these that can be used for educational purposes.

INTRODUCTION

Statistics is about collecting data, analysing it, figuring out what it means, and presenting it. The goal of statistics is to understand the data, not to use formulas to do multiple calculations.

Biostatistics: Biostatistics is the study of how statistics can be used in the biological or medical sciences. Francis Galton is thought to be the father of biostatistics. He came up with the statistical term "correlation." It is used in the fields of biology, medicine, nursing, pharmacy, public health and other health sciences to deal with statistics. Whether biostatistics is different from biometry depends on whether it is used in the health sciences (biostatistics) or in more general biology (biometry), such as agriculture, ecology, or wildlife biology. With the help of statistics, the biologist can find general laws from small samples and understand how variability works.

THE CENTRAL TENDENCY

It is a statistical metric that establishes a single value that properly characterizes the distribution's centre and represents the whole range of scores. This value can be represented by a single number.

Mean: The Arithmetic Mean is the result of dividing the total number of observations by the total number of observations.

Median: The variable's median positional value splits the distribution into two equal parts: the first half contains all of the values that are equal to or greater than the median value, while the second section includes all of the values that are equal to or smaller than the median value. It is possible to consider it the "middle" value for a given set of data.

Mode: When a concentration measurement is the most common in a series of observations, we say that it is in the mode. Mo represents the value that appears most often in a set of quantifiable values. To rephrase, it's the most often employed value in a given collection.

DISPERSION

The term "dispersion" is used to describe the way in which the objects diverge from both one another and from the mean. The more unlike the individual products in a series are to one another, the wider the dispersion will be. According to A.L. Bowley, the dispersion of the objects is a measurement of the variety of the items.

Standard deviation (σ): In 1983, Karl Pearson was the first person to propose the concept of a standard deviation. The standard deviation is the measure of dispersion among populations that is the most commonly used because of its practicality. It is represented by the Greek letter σ, which reads as "sigma". The value of each observation is factored into the calculation of the standard deviation, exactly as it is for the mean deviation.

Variance (σ^2): It shows how far apart all the data points in a set are.

STATISTICAL TOOLS

Microsoft excel: Excel is a popular statistics programme that can be used to look at the answers to hand-calculated homework problems and to learn about statistical concepts. It helps us to use Excel to do basic statistical analysis and present summaries of data. This covers basic data management tasks like tabulating data, making pivot tables and making graphics. It has been the most popular spreadsheet programme on this platform since version 5 came out in 1993. Excel is part of the Microsoft Office suite.

SPSS (Statistical Package for Social Sciences): With just one click, the highly interactive software SPSS can manipulate and analyse huge amounts of data in ways that would be impossible to do by hand. Researchers of all kinds use it to look at statistical data in a sophisticated way. The SPSS software package was made so that data from the social sciences could be managed and analysed statistically. It was made by IBM employees Norman H. Nie and C. Hadlai Hulll, and SPSS Inc. put it on the market in 1968. Later, in 2009, IBM bought SPSS Inc. SPSS can take data from almost any file type and use it to make tabulated reports, charts and plots of distributions and trends, summarise statistics and do a lot of statistical analysis.

Minitab: In 1972, researchers Barbara F. Ryan, Thomas A. Ryan, Jr., and Brian L. Joiner at Pennsylvania State University in the United States made Minitab, a powerful statistical programme. It began as a lighter version of OMNITAB 80. Most of what Minitab is used for is research and statistical analysis. Using Minitab for statistical analyses makes the

analysis more accurate, the results more reliable and the process go faster. It helps with exploratory data analysis by using sophisticated graphs, charts and other tools. The menu system is helpful for most analyses. It can open many different types of files such as text files, HTML files and Excel worksheets.

R : R is a piece of open-source software and a programming language that was initially developed in 1993 by Ross Ihaka and Robert Gentleman of the University of Auckland in New Zealand. R is a programming language that is often used by statisticians for the development of statistical software as well as data analysis. The S programming language was used in its creation.

QUESTION

The ages and weights of the first ten patients who visited the outpatient department (OPD) of a hospital on Monday are mentioned below. Calculate the mean, median, mode, and standard deviation and variance of the data., as well as Karl Pearson's correlation coefficient, using the information provided.

Table 1: The ages and weights of the 10 patients visited on Monday in a hospital OPD.

Patient ID	Age (In years)	Weight (In kg)
1	72	67
2	45	16
3	76	100
4	23	31
5	90	63
6	72	85
7	56	67
8	76	5
9	78	98
10	12	86
Mean	?	?
Median	?	?
Mode	?	?
SD	?	?
Variance	?	?

SOLUTION

MICROSOFT EXCEL

Calculation of mean

i. Launch Excel and enter the data into the rows and columns provided.
ii. Align the cells accordingly.
iii. Write the MEAN at the end of the patient ID column.
iv. In the next box, enter =AVERAGE(select all data from the age column) and hit enter. The determined mean age will be displayed in this box.
v. Choose this box and move the pointer to the next available weight column box.
vi. The calculated mean weight will be displayed in this box.
vii. Mean = 60 years (Age) & 61.8 Kg (Weight)

Calculation of median

i. Open Excel and enter the provided data into the rows and columns.
ii. Align the cells if necessary.
iii. At the bottom of the patient ID column, write MEDIAN.
iv. In the following box, type =MEDIAN(select all of the values in the age column) and then press the enter key on the keyboard. In that box, the calculated age median will be displayed.
v. Select that box and move the point to the next empty box in the weight column.
vi. The calculated weight median will be displayed in that box.
vii. Median = 72 years of age and 67 kg of weight

Calculation of mode

i. Enter data into an Excel spreadsheet.
ii. Align cells.
iii. Write MODE in the box below median under patient ID.
iv. Choose the next box and type =MODE(select all age column values) and press enter. That box will show the calculated age mode.
v. Choose that box and move to the next empty weight column box.
vi. The box displays the computed weight mode.
vii. Mode = 72 years (Age) and 67 kg (Weight)

Calculation of standard deviation (SD/σ)

I. Start Excel and enter the data.
ii. Align cells if needed.
iii. Add SD to the bottom patient ID coloumn below mode.
iv. In the next box, type =STDEVA(select all the age column data) and press Enter.
v. That box will display age standard deviation.
vi. Move the point to the weight column's next empty box from that box.
vii. SD = 25.69 years (Age) and 33.70 kg (Weight).

Calculation of Variance (σ^2)

i. Open Excel and enter the provided data into the rows and columns.
ii. Align the cells if necessary.
iii. At the bottom, add VARIANCE to the patient ID column.
iv. In the next box, type =VAR(select everything in the age column) and press the Enter key on the keyboard.
v. The calculated age variance will be displayed in that box.
vi. Select that box and move the point to the next empty box in the weight column.
viii. The calculated weight variance will be displayed in that box.
ix. 659.78 years (Age) and 1135.73 kg (Weight) are the variance.

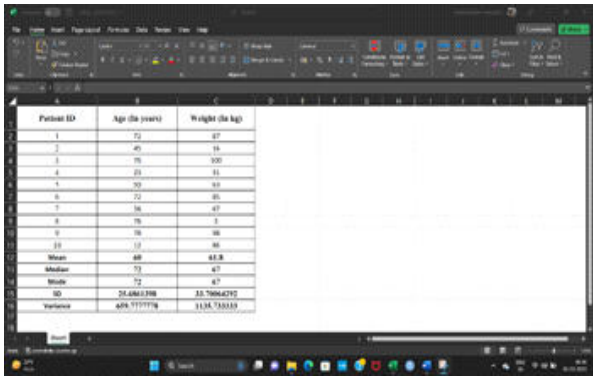

Fig.1: Calculation of mean, median, mode, standard deviation and variance using excel spreadsheet

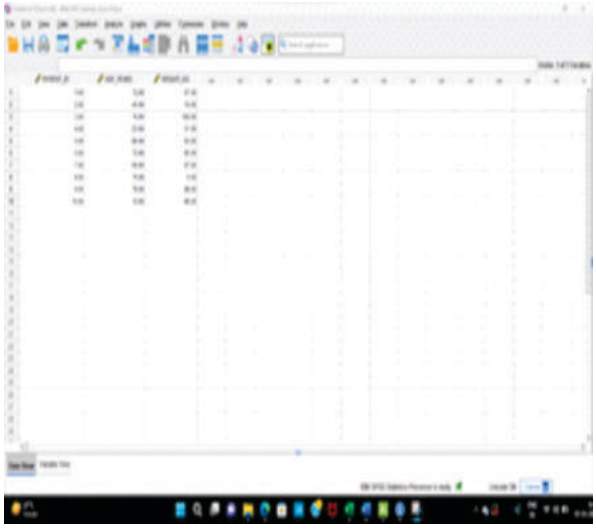

Fig. 2: Shows data of age and weight inserted in SPSS data view sheet

SPSS (STATISTICAL PACKAGE FOR SOCIAL SCIENCES):

It is a very interactive software.

I. Open the IBM SPSS statistics data editor worksheet.

ii. Feed all data in the data view sheet.

iii. Click on analyze.

iv. Choose descriptive statistics.

v. Then choose frequencies.

vi. A dialogue box will display in which select age and weight to the variables i.e. from left side box to right side box.

vii. Mark √ on display frequency tables.

viii. Click on statistics.

ix. Choose what quantities you need i.e. mean, median, mode, standard deviation and variance.

x. Click on continue then press OK.

Fig. 3: Shows the results of mean, median, mode, standard deviation and variance obtained in SPSS.

MINITAB

I. Open Minitab.

ii. Insert all data in columns.

iii. Click on stat.

iv. Select basic statistics.

v. Select display descriptive statistics.

vi. Choose variables i.e., age and weight.

vii. Click on statistics.

viii. Choose options i.e. mean, median, mode, standard deviation and variance.

ix. Press OK.

x. Press OK again.

xi. Calculated statistics output will display.

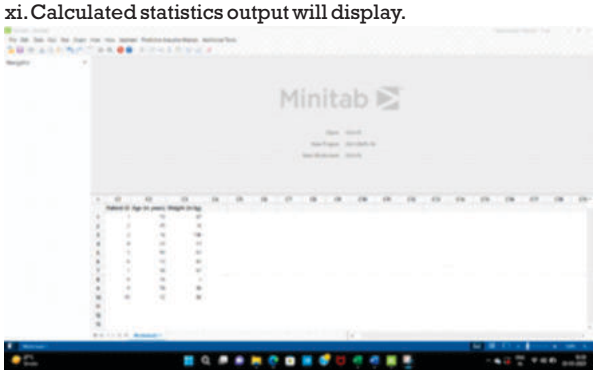

Fig. 4: Shows the data of age and weight inserted to calculate the central tendency and dispersion.

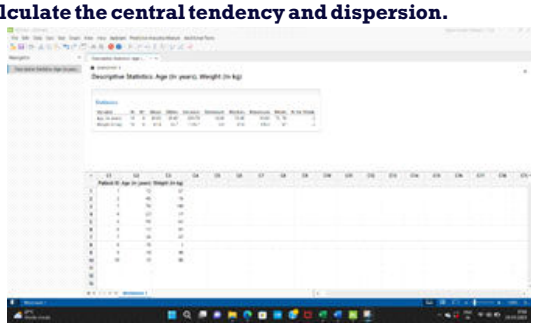

Fig. 5: Shows the output in Minitab for mean, median, mode, standard deviation and variance.

CONCLUSION

Due to the numerous benefits they offer, such as simplicity, less labour and time savings, the use of technologies is expanding at an ever-increasing rate. If a person is aware of how to manage and use software, he/she will be able to complete the task quickly and accurately. Biostatistics is a vital component of research and is not possible without it. The phrases mean, median and mode are frequently used while attempting to characterise central tendency. The terms standard deviation and variance are frequently used to describe attempts to assess dispersion. Pen and paper calculations require a significant amount of work and time to complete. Despite the fact that there are several statistical programmes accessible, including SPSS, R Online and Minitab, Microsoft Excel is by far the most common choice. This is because Excel is free and simple to use, which allows users to save time and money although it works with limitations. Apart for Microsoft Excel, no software is offered for free; trial versions are, however, available for a brief length of time. This article explains how to use MS Excel, SPSS, and Minitab to compute central tendency (mean, median and mode) and dispersion (standard deviation and variance). Definitely this article will be useful to both researchers and students who need to calculate central tendency and dispersion using MS Excel, SPSS, and Minitab because step by step procedure is given for each software.

REFERENCES:

[1] Singh, G., and Mahato, T.K., "Biostatistics and Research Methodology.", Technical publications, Pune, 1st ed., 2022.